

# Faktorianalyysi ja survey-aineiston tiivistäminen

**Kimmo Vehkalahti**

yliopistonlehtori, valtiot.toht.  
soveltavan tilastotieteen dosentti  
Opettajien akatemian jäsen

Yhteiskuntatilastotiede, Sosiaalitieteiden laitos  
Valtiotieteellinen tiedekunta, Helsingin yliopisto

[www.helsinki.fi/people/Kimmo.Vehkalahti/suomeksi.html](http://www.helsinki.fi/people/Kimmo.Vehkalahti/suomeksi.html)

**Sosiaalitutkimuksen tilastolliset menetelmät**  
Osa 2



# Osan 2 osaamistavoitteet ja sisältö

Osassa 2 tavoitteena on oppia **analysoimaan kyselytutkimus- eli survey-aineistoa faktorianalyysillä** sekä sen pohjalta **tiivistämään aineistoa** jatkoanalyysia varten.

Tällaiset aineistot ovat tyypillisiä mm. yhteiskuntatieteellisessä tutkimuksessa, jossa mielenkiinto kohdistuu esim. *asenteisiin*, *arvoihin* ja *mielipiteisiin*. Näitä **moniulotteisia ilmiöitä** mitataan kysely- ja haastattelulomakkeilla.

## 1. Faktorianalyysin perusteet

- ▶ Mittaaminen, mittausmalli ja korrelaatiot
- ▶ Faktorianalyysin merkitys ja historia
- ▶ Analyysin soveltaminen ja tulkinta

## 2. Survey-aineiston tiivistäminen

- ▶ Mittausmallista aineiston tiivistämiseen
- ▶ Faktoripiste- ja summamuuttujat



# 1. Faktoriansalyysin perusteet

**Tutkimuskysymyksistä johdettu mittausmalli** ohjaa sekä kyselylomakkeen laatimista että lomakkeen pohjalta kerätyn survey-aineiston analysointia.

- ▶ **mittausmalli:** **mitä** mitataan, **millä** ja **miten**
- ▶ mittausinstrumentti: **millä** mitataan ja **miten**
- ▶ kysely- ja haastattelututkimuksen instrumentti: **lomake**
- ▶ **mittari:** kokoelma saman aihepiirin mittauksia

## Mittauksen laatu:

1. **validiteetti:** mitataanko sitä mitä pitikin?
2. **reliabiliteetti:** mitataanko riittävän tarkasti?

## Mittautaso

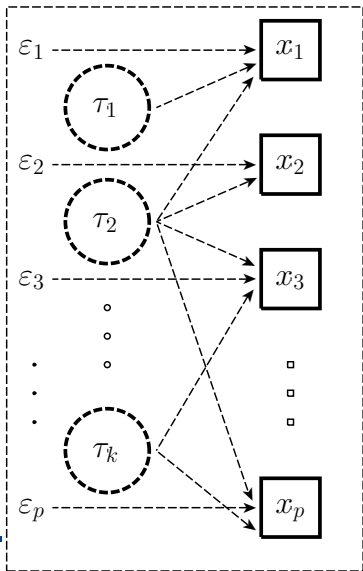
 vaikuttaa menetelmävalikoimaan:

- ▶ luokittelu — järjestäminen — numeerinen mittaus
- ▶ useimmat menetelmät edellyttävät numeerista mittautusta
- ▶ joissakin menetelmissä luokittelukin voi riittää
- ▶ järjestystasoinen mittaus usein ongelmallisinta



# Mittausmalli, johon faktorianalyysi perustuu

(vrt. Johdatus yhteiskuntatilastotieteeseen, Teema 1)



## Mittausmalliin sisältyy:

1. **faktoreita**  $\tau$  ( $k$  kpl), jotka vastaavat tutkittavan ilmiön ulottuvuuksia
2. **muuttujia**  $x$  ( $p$  kpl), jotka on mitattu; vastaavat kyselylomakkeen osioita
3. **mittausvirheitä**  $\varepsilon$  ( $p$  kpl), joita ei voi koskaan kokonaan välttää

# Mittausmalli, tutkimuskysymykset ja korrelaatiot

Paras käsitys faktoreista ja niiden lukumäärästä sekä malliin soveltuvista muuttujista pohjautuu **tutkimuskysymyksiin** ja **tutkimusalan teoriaan**. Lukumäärälle pätee:  $k < p$  eli faktoreita on aina vähemmän kuin mitattuja muuttujia.

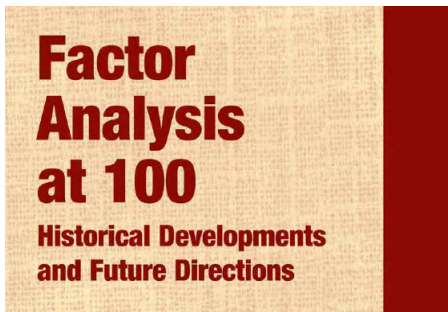
Faktoreiden ja muuttujien välisillä nuolilla osoitetaan niiden välisiä yhteyksiä, joita kuvataan **korrelaatioiden** avulla. *Nuolet: faktorin ajatellaan vaikuttavan siihen, miten kysymykseen vastataan.*

Kaikki katkoviivoilla esitetty on *hypoteettista*, koska aineisto koostuu vain mitatuista muuttujista. Tavoite on tiivistää aineistoa muodostamalla uusia muuttujia, jotka vastaavat mielenkiinnon kohteena olevia (hypoteettisia) faktoreita.



# Faktorianalyysin merkitys yhteiskuntatieteille

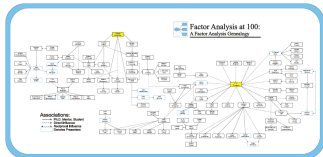
"Factor analysis is one of the great success stories of statistics in the social sciences because the primary focus of attention has always been on the relationships among fundamental traits such as intelligence, social class, or health status, that are *unmeasurable*."



Cudeck, Robert & MacCallum, Robert C., eds. (2007). Factor Analysis at 100: Historical Developments and Future Directions. [www.fa100.info](http://www.fa100.info)



# Faktorianalyysin ensimmäiset 111 vuotta (1904–2015)



Kehittäjien yhteydet 2004  
("akateeminen sukupuu"):

[www.fa100.info/tree050504.pdf](http://www.fa100.info/tree050504.pdf)

Kolme keskeisintä kehittäjää:

1. **C. E. Spearman** (1863–1945)
  - ▶ brittiläinen psykologi
  - ▶ julkaisuja faktorianalyysistä 1904–1927
2. **L. L. Thurstone** (1887–1955)
  - ▶ amerikkalainen psykologi
  - ▶ julkaisuja faktorianalyysistä 1931–1947
3. **K. G. Jöreskog** (s.1935)
  - ▶ ruotsalainen tilastotieteilijä
  - ▶ julkaisuja faktorianalyysistä 1967–



# Faktoriansalyysin ensimmäiset 111 vuotta (1904–2015)

Sadan vuoden aikajana (2004):

[www.fa100.info/timeline050504.pdf](http://www.fa100.info/timeline050504.pdf)



Viisi vaikuttavinta virstanpylvästä:

1. **Faktoriansalyysin varhainen synty: Spearman** (1904)
2. **Yleistys monikulotteisiin asetelmiin: Thurstone** (1931)
  - ▶ yksinkertainen faktorirakenne ("simple structure")
  - ▶ ortogonaalinen ja vino, **graafinen** faktorirotaatio
3. **Tilastotieteilijöiden mukaantulo:** Lawley (1940) ym.
4. **Ad hoc -metodit, analyttiset rotaatiot yms.:**
  - ▶ "*non-graphical rotation method*" (Horst 1941)
  - ▶ Transformaatioanalyysi (Ahmavaara 1954)
  - ▶ Quartimax (Neuhaus & Wrigley 1954)
  - ▶ Varimax (Kaiser 1958)
  - ▶ Procrustes (Hurley & Cattell 1962)
  - ▶ Oblimin (Jennrich & Sampson 1966)
  - ▶ "Target" (Schönemann 1966, Cliff 1966, Mustonen 1966)
5. **Konfirmatorinen analyysi, rakenneyhtälömallit: Jöreskog** (1967, 1969, 1970, 1973, 1979, ...) *et al*, Bollen (1989), ...





# 1. Faktorianalyysin varhainen synty (1904)

Kaksi *American Journal of Psychology* -lehden samassa numerossa ilmestynyttä artikkelia (30+93=123 sivua!) loivat (lähes) tyhjästä perustan **latenttien muuttujien mittaamiselle ja analysoinnille**.

## THE PROOF AND MEASUREMENT OF ASSOCIATION BETWEEN TWO THINGS.

By C. SPEARMAN.

Ensimmäinen artikkeleista (1904a) käsittelee **mittausvirheiden vaikutusta korrelaatioon** ja esittää *attenuaatiokaavan*, jolla tuota (korrelaatiota heikentävää) vaikutusta voidaan korjata.

Spearman, C. (1904a). The proof and measurement of association between two things. *American Journal of Psychology*, **15**, 72–101.



# 1. Faktorianalyysin varhainen synty (1904)

Jälkimmäinen artikkeli (1904b) koskee **ihmisen älykkyyttä** ja esittää teorian *yleisestä älykkyyksifaktorista (G)* sekä menetelmän, jolla tällaista faktoria voidaan tarkastella matemaattisesti.

“GENERAL INTELLIGENCE,” OBJECTIVELY  
DETERMINED AND MEASURED.

By C. SPEARMAN.

Spearman, C. (1904b). General intelligence objectively determined and measured. *American Journal of Psychology*, **15**, 201–293.

Tästä sai alkunsa (ja nimensä) **faktorianalyysi**, joka on kiistatta yksi **tilastotieteen menestystarinoista**, erityisesti yhteiskunta- ja käyttäytymistieteissä, mutta myös millä tahansa muillakin aloilla, joissa tutkimuksen kohteina ovat erilaiset latentit ilmiöt.



## 2. Yleistys monikulotteisiin asetelmiin (1930-luku)

Thurstone, L. L. (1931). Multiple Factor Analysis. *Psychological Review*, **38**, 406–427.

- ▶ yksinkertainen faktorirakenne ("simple structure")
- ▶ ortogonaalinen ja vino, **graafinen** faktorirotaatio  
[www.youtube.com/watch?v=-yZRUR6gpAg](http://www.youtube.com/watch?v=-yZRUR6gpAg) (Survo-demo)
- ▶ famous box problem  
[www.youtube.com/watch?v=76AELq81DQs](http://www.youtube.com/watch?v=76AELq81DQs) (Survo-demo)

- ▶ "*Psychology as a Quantitative Rational Science*"

Thurstonen kuuluisa, Psykometriikan seuran (*Psychometric Society*) perustamiskokouksessa 1936 pitämä puhe, joka julkaistiin näkyvyyden takaamiseksi *Science*-lehdessä:

Thurstone, L. L. (1937). Psychology as a quantitative rational science. *Science*, **85**, 227–232.



### 3. Tilastotieteilijöiden mukaantulo (1940–1950–1960)

Faktorianalyysi sai siis alkunsa psykologian ja **psykometriikan** puolella. Tilastotieteilijät ”hyväksyivät” sen tilastolliseksi menetelmäksi vasta paljon myöhemmin – faktorianalyysi oli jo sitä ennen vuosikymmeniä aikaansa edellä.

Lawley, D. N. (1940). The estimation of factor loadings by the method of maximum likelihood. *Proceedings of The Royal Society of Edinburgh*, **60**, 64–82.

Rao, C. R. (1955). Estimation and tests of significance in factor analysis. *Psychometrika*, **20**, 93–111.

Anderson, T. W. & Rubin, H. (1956). *Statistical Inference in Factor Analysis*. Proceedings of the Third Berkeley Symposium on Mathematical Statistics and Probability, Volume 5, 111–150, University of California Press, Berkeley, CA.  
<http://projecteuclid.org/euclid.bsm/1200511860>

Lawley, D. N., & Maxwell, A. E. (1963). *Factor Analysis as a Statistical Method*. Butterworths, London. (2. painos 1971).

■ **Viimeksi mainitun teoksen nimi on paljon puhuva!**



## 4. Ad hoc -metodit, analyttiset rotaatiot yms.

(Näihin vaiheisiin ei tässä yhteydessä mennä yksityiskohtaisemmin.)

Faktorianalyysia päästiin kunnolla soveltamaan vasta tietokoneiden tullessa hiljalleen käyttöön 1960-luvulta lähtien. Tällöin saivat alkunsa monet tilastolliset ohjelmistot kuten Survo, SPSS ja SAS.

Laskentatehon kasvua odotellessa oli ehditty kehittää *ad hoc* -laskenta-algoritmeja ja "kikkakonsteja", joista osa on sitkeästi juuttunut käyttöön, vaikka parempia vaihtoehtoja olisi olemassa.

*"Factor analysis had to mark time until technology caught up."*

Bartholomew, D. J. (1995). Spearman and the origin and development of factor analysis. *The British Journal of Mathematical and Statistical Psychology*, **48**, 211–220.



## 5. Konfirmatorinen analyysi, rakenneyhtälömallit

Uppsalan yliopiston professori *Karl Gustaf Jöreskog* onnistui väitöskirjassaan (1967) paikkaamaan laskenta-algoritmeja häirinneen ongelman, joka esti suurimman uskottavuuden menetelmän soveltamisen faktorianalyysin parametrien estimoinnissa.

Jöreskogin ansiosta menetelmää voitiin vihdoinkin soveltaa käytännön tutkimuksissa käyttäen tilastollisesti pätevää menetelmää.

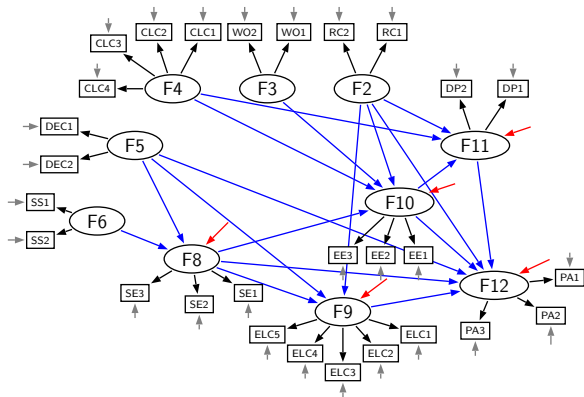
Jöreskog, K. G. (1967). Some contributions to maximum likelihood factor analysis. *Psychometrika*, **32**, 443–482.

Tästä lähti liikkeelle laaja teoriakehitys. Se johti seuraavina vuosina (ja vuosikymmeninä) **konfirmatorisen faktorianalyysin** syntyyn ja **rakenneyhtälömalleihin**, jotka yhdistävät faktorianalyysin, regressioanalyysin ja polkumallit.



## 5. Konfirmatorinen analyysi, rakenneyhtälömallit

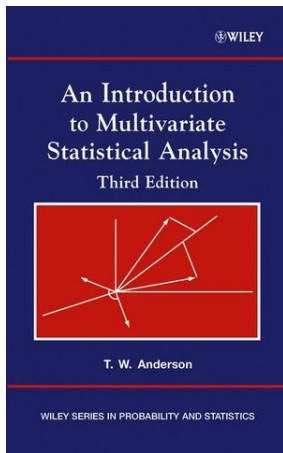
Rakenneyhtälömalleja (*Structural Equation Models, SEM*) sovelletaan nykyisin paljon monenlaisissa, erityisesti yhteiskunta- ja käyttäytymistieteellisissä tutkimusasetelmissä.



# Monimuuttujamenetelmien kehittyminen

**Monimuuttujamenetelmät** kehittyivät moniulotteisen faktorianalyysin tapaan 1930–1940 -luvulta lähtien.

Keskeiseksi, kokoavaksi teokseksi muodostui varhain T.W. Andersonin (1958) teos *An Introduction to Multivariate Statistical Analysis*, josta sittemmin ilmestyi 2. painos (1984) ja 3. painos (2003).



**Tietokoneiden ja tilastollisten laskentaohjelmistojen myötä monimuuttujamenetelmät yleistyivät nopeasti yhä useammilla tieteenaloilla.**





# Faktoriansalyysin suomalainen koulukunta

Suomessa monimuuttujamenetelmät tulivat tutuiksi prof. Seppo Mustosen 1960-luvulla pitämällä kursseilla, jotka perustuivat aluksi em. Andersonin kirjaan (ks. tarkemmin [www.survo.fi/monim](http://www.survo.fi/monim)).

Kiinnostava katsaus tilanteeseen **50 vuotta sitten**:

Mustonen, Seppo (1965). *Tilastollisista monimuuttujamenetelmistä*. Alkoholipoliittisen tutkimuslaitoksen tutkimusseloste N:o 14. [www.survo.fi/publications/Tilastollisista\\_monimuuttujamenetelmista\\_Mustonen\\_1965.pdf](http://www.survo.fi/publications/Tilastollisista_monimuuttujamenetelmista_Mustonen_1965.pdf)

Seppo Mustosen oppikirja *Tilastolliset monimuuttujamenetelmät* (1995) on yhä (20 v ilmestymisensä jälkeen) paras suomenkielinen perusteos faktoriansalyysin ja muiden keskeisten menetelmien teoriasta ja sovelluksista.

Koko kirja on luettavissa verkossa:  
■ [www.survo.fi/mustonen/monim.pdf](http://www.survo.fi/mustonen/monim.pdf)



# Faktorianalyysin soveltaminen ja tulkinta

Perehdytään nyt faktorianalyysin soveltamiseen ja tulkintaan tyypillisessä kyselytutkimusasetelmassa. Kiinnostuksen kohteena olevat ilmiöt ovat asenteita, arvoja tms., joita ei voi suoraan havaita tai mitata. On käytettävä epäsuoria mittaustapoja kuten kyselylomaketta, joka koostuu useista, ko. ilmiöitä mittaavista osioista eli kysymyksistä tai väitteistä.

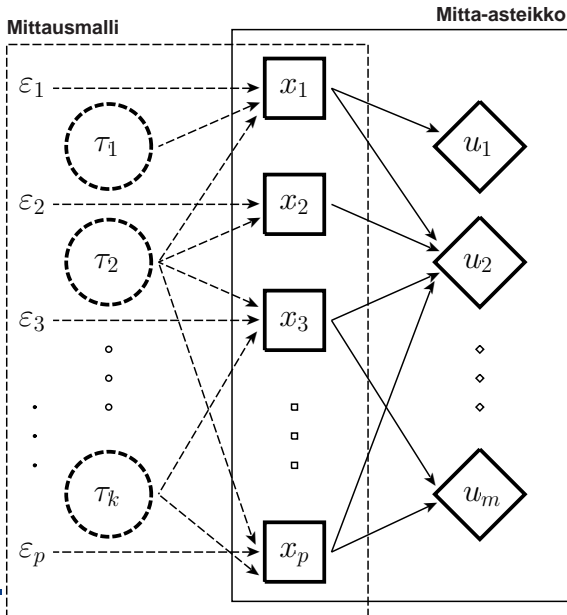
Faktorianalyysissa on kaksi vaihetta: **vaiheen 1** tavoitteena on hahmottaa aineiston **taustalla oleva rakenne**, jota voidaan etukäteen kuvata mittausmallilla. Lähestymistapoja ovat *eksploratiivinen* eli aineistoperustainen ja *konfirmatorinen* eli malliperustainen faktorianalyysi. Käytännön analyysit ovat näiden väliltä eli niissä on tyypillisesti piirteitä molemmista.

Faktorianalyysin **vaiheen 2** myötä **aineistoa tiivistetään** vaiheen 1 tulosten ja tutkimusalan teorian tuella.

■ Enemmän tässä esitetyistä asioista löytyy mm. oppikirjani **Kyselytutkimuksen mittarit ja menetelmät** luvuista 4–7.



# Mittausmallista aineiston tiivistämiseen (faktorianalyysi)



# Mittausmalli ja mitta-asteikko

## Mittausmalli (vaiheessa 1 fokus on tässä!)

- ▶ Mitä **ilmiötä** tutkitaan? Montako **ulottuvuutta** siinä on?
- ▶ Millä ilmiötä **mitataan** — mahdollisimman hyvin?
- ▶ Yhteyksien ja ulottuvuuksien tutkiminen: **faktorianalyysi**
- ▶ Esimerkki: ESS (European Social Survey)

## Mitta-asteikko

- ▶ Osioiden eli mitattujen muuttujien yhdistelmä
- ▶ Yhdistelytapoja: **faktoripisteet**, **summamuuttujat**
- ▶ Yleinen tavoite: **aineiston tiivistäminen**
- ▶ Esimerkki: ESS (European Social Survey)



# Esimerkki: faktorianalyysi (ESS)

ESS (European Social Survey), round 5, 2010, Suomi (n=1878)

Tarkastellaan tässä vain seuraavia muuttujia:

- ▶ How interested in politics (**1=very**, ..., 4=not at all)

Seuraavissa asteikko 0=no trust at all, ..., 10=complete trust:

- ▶ Trust in country's parliament
- ▶ Trust in the legal system
- ▶ Trust in the police
- ▶ Trust in politicians
- ▶ Trust in political parties
- ▶ Trust in the European Parliament
- ▶ Trust in the United Nations

Seuraavissa asteikko 0–10 eri sanamuodoin (10=positiivisin):

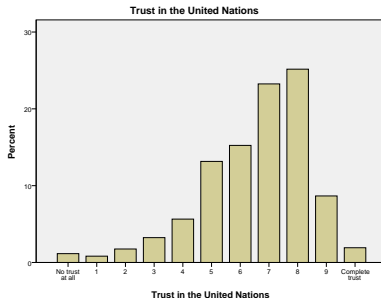
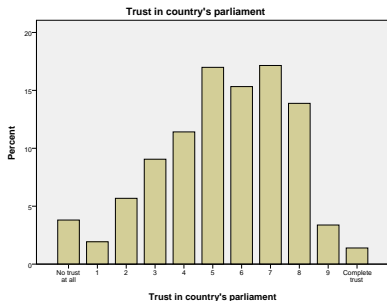
- ▶ Most people can be trusted or you can't be too careful
- ▶ Most people try to take advantage of you, or try to be fair
- ▶ Most of the time people are helpful or mostly looking out for themselves

(Kysymysten tarkemmat sanamuodot on tässä sivuutettu.)



# Esimerkki: faktorianalyysin oletukset

Muuttujien jakaumat vaihtelevat, olennaista on mittausaste:



Oletetaan nyt, että olisi hahmoteltu **3 faktorin** mittausmalli (ulottuvuuksina mm. luottamus poliittisiin toimijoihin ja ihmisiin).

Tehdään seuraavaksi tältä pohjalta faktorianalyysi SPSS:llä.

*(Esimerkkitalanne on tarkoituksella valittu suppeaksi ja osin keinoiseksi, koska fokus on enemmän menetelmässä ja sen*

- *tulkinnoissa. Aineisto mahdollistaa sisällöllisesti paljon monipuolisempiakin asetelmia!*

# Esimerkki: faktorianalyysin suoritus SPSS:llä

Faktorianalyysi tiivistää tiedot muuttujien välisistä korrelaatioista annetuksi määräksi faktoreita.

\* faktorianalyysi 3-ulotteisen mittausmallin pohjalta (SPSS):

FACTOR

/VARIABLES polintr trstprl trstlgl trstplc trstplt trstprt trstep  
trstun ppltrst pplfair pplhlp

/MISSING LISTWISE

/ANALYSIS polintr trstprl trstlgl trstplc trstplt trstprt trstep trstun  
ppltrst pplfair pplhlp

/PRINT INITIAL EXTRACTION ROTATION

/FORMAT SORT

/CRITERIA **FACTORS(3)** ITERATE(25)

/EXTRACTION ML

/CRITERIA ITERATE(25)

/ROTATION VARIMAX.



# Esimerkki: faktorimatriisi ja faktorilataukset

Faktorianalyysin olennaisin tuloste on ns. rotatoitu faktorimatriisi:

Rotated Factor Matrix<sup>a</sup>

	Factor		
	1	2	3
Trust in politicians	,895	,249	,140
Trust in political parties	,876	,237	,135
Trust in country's parliament	,742	,205	,301
Trust in the European Parliament	,703	,187	,213
Trust in the United Nations	,452	,210	,368
How interested in politics	-,212	-,005	-,088
Most people can be trusted or you can't be too careful	,160	,720	,183
Most people try to take advantage of you, or try to be fair	,099	,686	,153
Most of the time people helpful or mostly looking out for themselves	,195	,582	,104
Trust in the legal system	,378	,214	,745
Trust in the police	,177	,206	,651

Extraction Method: Maximum Likelihood.  
Rotation Method: Varimax with Kaiser Normalization.

- (järjestettynä faktorilatausten perusteella faktoreittain)



# Esimerkki: kommunaliteetit

On syytä tarkastella myös muuttujien kommunaliteetteja:

Communalities

	Initial	Extraction
How interested in politics	,055	,053
Trust in country's parliament	,656	,683
Trust in the legal system	,538	,743
Trust in the police	,392	,498
Trust in politicians	,801	,883
Trust in political parties	,776	,841
Trust in the European Parliament	,619	,575
Trust in the United Nations	,459	,383
Most people can be trusted or you can't be too careful	,397	,577
Most people try to take advantage of you, or try to be fair	,345	,504
Most of the time people helpful or mostly looking out for themselves	,299	,387

Extraction Method: Maximum Likelihood.

■ (valitettavasti SPSS antaa ne eri järjestyksessä eri taulukkoon!)

# Esimerkki: tulosten esittäminen ja tulkinta

Factor structure of ESS (Finland), $n = 1878$	F1	F2	F3	$h^2$
<b>F1: Trust in Political Actors</b>				
Trust in politicians	<b>0.90</b>	0.25	0.14	0.88
Trust in political parties	<b>0.88</b>	0.24	0.14	0.84
Trust in country's parliament	<b>0.74</b>	0.21	<i>0.30</i>	0.68
Trust in the European Parliament	<b>0.70</b>	0.19	0.21	0.58
Trust in the United Nations	<i>0.45</i>	0.21	<i>0.37</i>	0.38
How interested in politics	-0.21	-0.01	-0.09	0.05
<b>F2: Trust in People</b>				
Most people can be trusted	0.16	<b>0.72</b>	0.18	0.58
Most people try to take advantage of you	0.10	<b>0.69</b>	0.15	0.50
Most of the time people are helpful	0.20	<b>0.58</b>	0.10	0.39
<b>F3: Trust in Authorities</b>				
Trust in the legal system	<i>0.38</i>	0.21	<b>0.75</b>	0.74
Trust in the police	0.18	0.21	<b>0.65</b>	0.50
Sum of squares	3.11	1.66	1.36	6.13
Variance explained %	28.3	15.0	12.4	55.7

$h^2$  = communalities

# Esimerkki: faktorianalyysin tulkinnoista

(*Taulukoita ja tulkintoja katsastellaan tarkemmin luennolla.*)

Edeltä tulostaulukoista ilmenee, että kolme faktoria ”selittää” yhteensä 55.7 % kyseisten muuttujien välisestä vaihtelusta.

Olenneisempia ovat faktorien ja muuttujien väliset yhteydet (korrelaatiot), joita tulkitaan faktorilatausten avulla (*vrt. nuolet*).

Ensimmäinen faktori (”luottamus poliittisiin toimijoihin”?) on voimakkain. Toinen voisi olla nimeltään ”luottamus ihmisiin” ja kolmas ”luottamus auktoriteetteihin”. Mitä enemmän nojataan tutkimusalan teoriaan, sitä helpompi on nimetä faktorit. (Teoria saattaa myös tukea *korreloivia faktoreita*; tässä ne on oletettu korreloimattomiksi, mikä on ainakin alkuvaiheessa selkeämpää.)

Kiinnostus politiikkaan, jolla on negatiivinen lataus (muista asteikon suunta!), ei nouse juurikaan esiin millään faktorilla. Samaa kuvastaa myös sen kommunaliteetti, joka on käytännössä nolla.



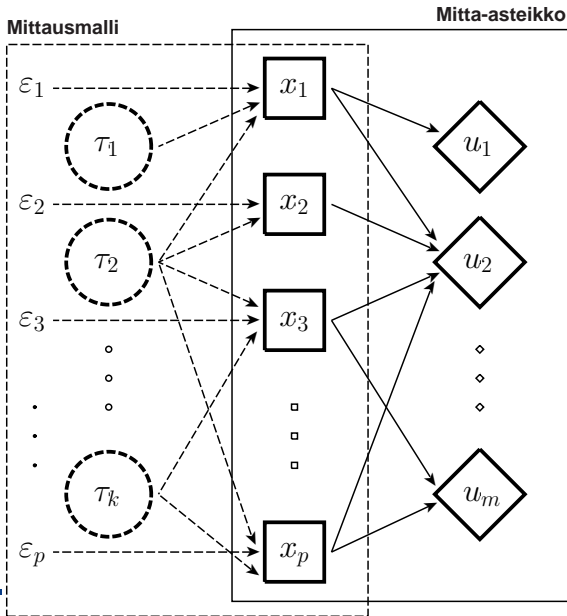
## 2. Survey-aineiston tiivistäminen

**Tilastollisten menetelmien** yleinen tavoite on **tiivistää** aineistoon sisältyvää informaatiota kuviksi, tunnusluvuiksi, taulukoiksi yms.

- ▶ tiivistäminen (ja muu analysointi) edellyttää ehdottomasti kunnollista aineistoon **tutustumista**
- ▶ tutustumisen ja tiivistämisen kannalta keskeistä: **jakaumien kuvaileminen** (graafisesti ja tunnusluvuin)
  - ▶ (empiirinen) **jakauma**: aineiston yhden muuttujan **kaikki** (mitatut ja koodatut) arvot
- ▶ tiivistäminen edellyttää kuitenkin yleensä enemmän:
  - ▶ kyselyaineistoissa on paljon ("liikaa") muuttujia
  - ▶ muuttujien **yhdistely** tutkimusalan teorian perusteella
  - ▶ **faktorianalyysi**: faktoripisteet, summamuuttujat
  - ▶ muut monimuuttuja- ym. analyysimenetelmät



# Mittausmallista aineiston tiivistämiseen (faktorianalyysi)





## Mittausmalli

- ▶ Mitä **ilmiötä** tutkitaan? Montako **ulottuvuutta** siinä on?
- ▶ Millä ilmiötä **mitataan** — mahdollisimman hyvin?
- ▶ Yhteyksien ja ulottuvuuksien tutkiminen: **faktorianalyysi**
- ▶ Esimerkki: ESS (European Social Survey)

## Mitta-asteikko (vaiheessa 2 fokus on tässä!)

- ▶ Osioiden eli mitattujen muuttujien yhdistelmä
- ▶ Yhdistelytapoja: **faktoripisteet**, **summamuuttujat**
- ▶ Yleinen tavoite: **aineiston tiivistäminen**
- ▶ Esimerkki: ESS (European Social Survey)

# Faktorianalyysin tulkinnasta aineiston tiivistämiseen

Faktorianalyysin toinen vaihe on aineiston **tiivistäminen** löydetyn faktorirakenteen perusteella. Siinä ”päästään eroon” suuresta määrästä muuttujia, joita on tarvittu ilmiön mittaamiseen, mutta joista on vaikea sellaisenaan saada kunnollista kokonaiskäsitystä.

Tiivistämisen onnistuminen edellyttää, että on löydetty oikea **faktorilukumäärä**, joka siis vastaa ilmiön ulottuvuuksien määrää. Tässä ennalta pohdittu mittausmalli on avainasemassa, sillä jos lukumäärä hahmotetaan vain aineiston pohjalta, jää turhan paljon epävarmuuksia. On tutkijan tehtävä päättää oikea lukumäärä.

Toinen edellytys on, että faktorit on pystytty **nimeämään** ja **tulkitsemaan** ymmärrettävästi. Myös tässä ilmiön tuntemus on ratkaisevan tärkeää. Ellei tulkinta onnistu, jää aineiston tiivistäminen keinotekoiseksi eikä siitä ole jatkossa vastaavaa hyötyä muiden analyysien tai visualisointien kannalta.



## Esimerkki: faktorianalyysi (ESS), jatkoa aiempaan

ESS (European Social Survey), round 5, 2010, Suomi (n=1878)

Tarkastellaan tässä vain seuraavia muuttujia:

- ▶ How interested in politics (**1=very**, ..., 4=not at all)

Seuraavissa asteikko 0=no trust at all, ..., 10=complete trust:

- ▶ Trust in country's parliament
- ▶ Trust in the legal system
- ▶ Trust in the police
- ▶ Trust in politicians
- ▶ Trust in political parties
- ▶ Trust in the European Parliament
- ▶ Trust in the United Nations

Seuraavissa asteikko 0–10 eri sanamuodoin (10=positiivisin):

- ▶ Most people can be trusted or you can't be too careful
- ▶ Most people try to take advantage of you, or try to be fair
- ▶ Most of the time people are helpful or mostly looking out for themselves

(Kysymysten tarkemmat sanamuodot on tässä sivuutettu.)





# Esimerkki faktorianalyysin tulosten esittämisestä

Factor structure of ESS (Finland), $n = 1878$	F1	F2	F3	$h^2$
<b>F1: Trust in Political Actors</b>				
Trust in politicians	<b>0.90</b>	0.25	0.14	0.88
Trust in political parties	<b>0.88</b>	0.24	0.14	0.84
Trust in country's parliament	<b>0.74</b>	0.21	<i>0.30</i>	0.68
Trust in the European Parliament	<b>0.70</b>	0.19	0.21	0.58
Trust in the United Nations	<i>0.45</i>	0.21	<i>0.37</i>	0.38
How interested in politics	-0.21	-0.01	-0.09	0.05
<b>F2: Trust in People</b>				
Most people can be trusted	0.16	<b>0.72</b>	0.18	0.58
Most people try to take advantage of you	0.10	<b>0.69</b>	0.15	0.50
Most of the time people are helpful	0.20	<b>0.58</b>	0.10	0.39
<b>F3: Trust in Authorities</b>				
Trust in the legal system	<i>0.38</i>	0.21	<b>0.75</b>	0.74
Trust in the police	0.18	0.21	<b>0.65</b>	0.50
Sum of squares	3.11	1.66	1.36	6.13
Variance explained %	28.3	15.0	12.4	55.7

$h^2$  = communalities



## Esimerkki: faktorianalyysin suoritus (jatkoa)

Faktorianalyysin toisessa vaiheessa tiivistetään aineiston tiedot uusiksi faktoripistemuuttujiksi.

\* faktorianalyysi 3-ulotteisen mittausmallin pohjalta (SPSS):

FACTOR

/VARIABLES polintr trstprl ...

... (samoin kuin aiemmin)

**/SAVE REG(ALL).**

\* nimetään uudet muuttujat faktoreiden mukaisesti:

RENAME VARIABLES

(FAC1\_1=Trust1)

(FAC2\_1=Trust2)

(FAC3\_1=Trust3).

VARIABLE LABELS

Trust1 'luottamus poliittisiin toimijoihin (fp)'

Trust2 'luottamus ihmisiin (fp)'

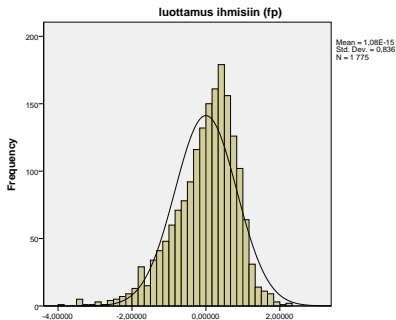
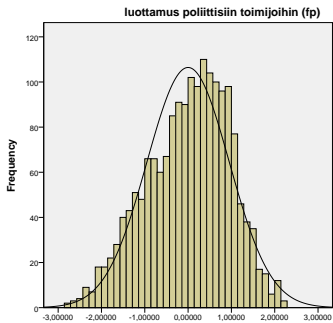
Trust3 'luottamus auktoriteetteihin (fp)'.



# Esimerkki: faktoripisteet ja niiden jakaumat

Kun faktorit on tulkittu ja nimetty, on aika siirtyä takaisin aineiston havaintoyksikötasolle. Muodostetut uudet faktoripistemuuttuja kertovat, mihin kohtaan mitäkin ulottuvuutta kuvaavaa jatkumoa kukin vastaaja sijoittuu.

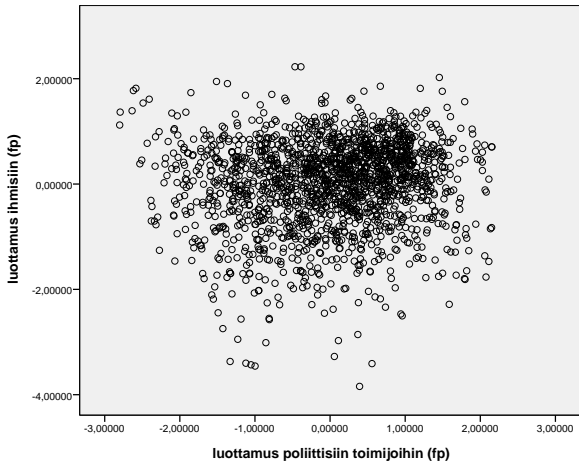
Visualisoidaan kahden faktoripistemuuttujan jakaumia:



■ Faktoripisteet ovat alkuperäisiä jatkuvampia (vinoutta on yhä).

# Esimerkki: faktoripisteiden väliset yhteydet

Visualisoidaan näiden samojen muuttujien yhteisjakaumaa:

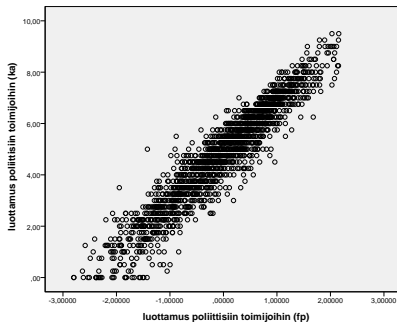
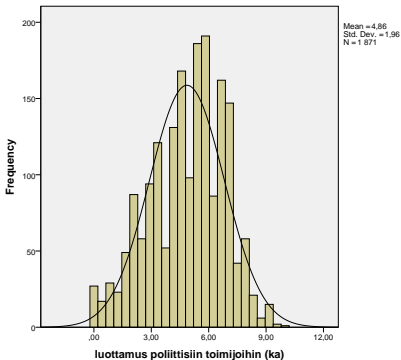


■ Hahmotatko "luottamus/epäluottamus"-nelikentän? Tulkitse!

# Esimerkki: summamuuttujien muodostaminen

Vaihtoehtoinen tapa on tehdä **summamuuttujia**, mutta ne eivät hyödynnä faktorianalysista saatua tietoa niin hyvin (esim. sitä, *miten hyvin* eri muuttujat mittaavat kutakin ulottuvuutta).

Histogrammi neljän ensimmäisen luottamusmuuttujan summamuuttujasta (keskiarvo) sekä hajontakuva siitä ja vastaavasta faktoripistemuuttujasta:



# Esimerkki: faktoripisteiden luokittelu

Faktoripisteet voidaan edelleen tiivistää esim. ala- ja yläkvartiilien avulla (<25 %, 25–75 %, >75 %) kolmeen luokkaan: 1="epäluottamus", 2="siltä väliltä", 3="luottamus".

Kahden tällaisen muuttujan välinen ristiintaulukointi:

luottamus poliittisiin toimijoihin (123) \* luottamus ihmisiin (123) Crosstabulation

Count		luottamus ihmisiin (123)			Total
		1 epäluottamus	2 siltä väliltä	3 luottamus	
luottamus poliittisiin toimijoihin (123)	1 epäluottamus	139	205	100	444
	2 siltä väliltä	218	461	208	887
	3 luottamus	88	220	136	444
Total		445	886	444	1775

Tämän tyyppiset tiivistykset ovat tyypillisiä yhteiskuntatieteissä. Jatkuvat muuttujat toimivat edellä vain välivaiheena. Informaatiota hukkuu (paljon), mutta kokonaisuus voi olla helpompi hahmottaa. (Monesti saatetaan tiivistää vain kahteen luokkaan!)



# Faktoripiste- ja summamuuttujat

Edellä on mainittu summamuuttujat faktoripisteiden vaihtoehtona. Kummankin käyttöön on perusteensa. Vertaillaan vähän:

**Faktoripisteet** muodostetaan faktorianalyysin perusteella, jolloin ne pyrkivät vastaamaan mahdollisimman hyvin aikaansaatuja faktoreita. Paremmat muuttujat huomioidaan suuremmilla painoilla kuin huonommat. Muuttujat voivat olla erilaisilla asteikoilla mitattuja sekä eri suuntaisia. Jos faktorit oletetaan keskenään korreloimattomiksi, faktoripisteetkään eivät korreloi keskenään.

**Summamuuttujat** muodostetaan valitsemalla vain parhaita muuttujia, mutta painottamalla niitä keskenään samanarvoisesti. Tutkimusalan teoria voi "sanella" valittavat muuttujat, jolloin ei faktorianalyysia periaatteessa tarvita lainkaan. On kuitenkin hyvä tarkistaa, miltä tilanne aineistossa näyttää. Muuttujien on joka tapauksessa oltava vertailukelpoisilla asteikoilla mitattuja sekä samansuuntaisia. Summamuuttujat korreloivat yleensä keskenään.



# Aineiston tiivistäminen: johtopäätöksiä

Aineiston tiivistäminen on tärkeä ja välttämätön vaihe, kun työskennellään kyselytutkimusaineiston parissa. Mittaus ei onnistu kovin vähillä muuttujilla, mutta kokonaiskäsitelysten saamiseksi muuttujia on aluksi ”liikaa”. Tiivistämisen pohjalta pitäisi saada hyvät mahdollisuudet jatkaa eteenpäin. On paljon helpompi jatkaa, kun yhtäaikaa käsiteltävien muuttujien määrä on pienempi.

**Monenlaiset kiinnostavat analyysit voivat alkaa vasta tästä!**

Faktoripisteet (tai summamuuttujat) voivat toimia esimerkiksi regressioanalyysin tai varianssianalyysin selitettävänä muuttujina (tai selittäjinä).

Jatkon kannalta olennaisen tärkeää on faktorianalyysin toteutus huolellisesti: hahmotetaan faktoreiden oikea (sopiva) lukumäärä sekä nimetään ja tulkitaan faktorit sisällöllisesti mielekkäästi.

